DOCUMENT RESUME

ED 156 717 TM 007 279

AUTHOR Rodriquez-Brown, Flora V.
TITLE The Do's and Don'ts in Regard to the Evaluation of Bilingual Programs.
PUB DATE Apr 78
NOTE 18p.; Paper presented at the Annual Meeting of the National Association for Bilingual Education (7th, San Juan, Puerto Rico, April 24-28, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Bilingual Education; *Control Groups; Data Collection; *Evaluation Methods; *Evaluation Needs; Evaluators; Experimental Groups; Longitudinal Studies; Predictor Variables; *Program Evaluation; *Research Design; Research Problems
IDENTIFIERS Quasi Experimental Design

ABSTRACT

The difficulties in evaluating bilingual education appear to have prevented success in all but a few evaluation attempts, but better and more meaningful evaluation is neccessary in order to identify the strengths and weaknesses of bilingual programs. Many bilingual programs are undergoing constant modification, adequate assessment instruments have been lacking, and political pressures have interfered with evaluation programs. Since it is hard to find individuals who are expert in both evaluation and bilingual education, evaluation teams are recommended rather than individual evaluators. Control groups are not normally available, so quasi-experimental designs must be used. The author recommends the Before-and-After and particularly the Time Series designs. Several precautions are listed. Independent variables for an evaluation are listed and categorized as contextual variables, student variables, and treatment variables. Dependent variables must reflect the goals or objectives of the bilingual program and must be valid. Good comprehensive documentation is also essential, particularly in longitudinal research. (CTM)

THE DO'S AND DON'TS IN REGARD TO THE EVALUATION OF BILINGUAL PROGRAMS

Flora V. Rodríguez-Brown, Ph.D.
Bilingual Education Service Center
Illinois Resource Center

(Paper presented at the Seventh Annual Meeting of the National Association for Bilingual Education in San Juan, Puerto Rico, April 24 - 28, 1978.)

Introduction

Bilingual education is enjoying its first decade of prominence in the United States. In 1963, Dade County in Florida started a public school Spanish-English bilingual program for Cuban Americans and Anglos. In 1967, the Bilingual Education Act was added to the Elementary and Secondary Education Act of 1965. Federally-funded Title VII bilingual education programs began in 1968. More recently, states have passed legislation to fund bilingual programs (Swanson, 1974; U.S. Commission on Civil Rights, 1975).

At a time when the implementation of bilingual programs has reached such a peak, the evaluation of programs lags far behind. Despite millions spent on the development of programs, the United States experience to date has yielded few meaningful insights into various aspects of program design (Troike, 1974; U.S. Commission on Civil Rights, 1975; Ramirez et al., 1975). Reasons for this lack of hard data include the following:

(1) It is hard, if not impossible, to obtain meaningful research results from pilot programs that are constantly undergoing modification, presumably for the better. Even if summative results are obtained, the researcher is hard put to give a label to a particular treatment, since it is in such a state of flux.

(2) There has been such a pressing need for formative evaluation of project-oriented goals, specifically behavioral objectives contained in the curriculum, that no time has remained for evaluating other things.

(3) Until recently there have not been adequate assessment instruments particularly bilingual ones, and even now much test development and norming are called for.

(4) Political threats to bilingual schooling have almost forced evaluation reports to be public relations documents.

(5) Evaluators have tended to be persons unfamiliar with the particular needs and characteristics of bilingual education.

The "fledgling program" reason should no longer apply, since bilingual projects nationwide now have more stability, as a result of a gradually growing accumulation of experience, methods, and material. But if bilingual education is to continue to advance, better and more meaningful evaluation is necessary.

With respect to a project-centered instructional objectives, more than ever before there is a need to entertain the larger questions as well. Tucker and d'Anglejan (1971) question whether "self centered" project goals such as meeting specific teaching objectives, are valid criteria for evaluating the success or failure of a program (e.g., 75% of the children can answer 90% of the questions in a certain section of a book). Whether or not such criteria are valid, there is more to formative evaluation, such as investigation of the following areas (adapted from Saville and Troike, 1971):

(1) The teaching techniques that prove most successful in different situations (grouping, sequencing and pacing of materials, and correction procedures).

(2) The effect of program design (e.g., partial or full bilingual schooling using a concurrent, dual language, or alternate days approach to instruction).

(3) The effect of teacher training and patterns of staff utilization.

The lack of adequate instruments is still a problem, though not as severe as 10 years ago when federally-funded program evaluations were initiated. Yet evaluation must proceed even if the most appropriate instrument is not available. Recently even some widely used standardized instruments, such as the Peabody Picture Vocabulary Test and the Cooperative Primary Test of Reading, have been subject to criticism (Cicourel et al., 1974). We seem to be entering an era in which ethnomethodological scrutiny of tests and of individual test items will be common practice.

Finally, it would appear that bilingual schooling is here to stay, at least for the foreseeable future. Thus, evaluation reports should reflect more than a morass of tabular data and a scattering of carefully selected and tentatively--or even ambiguously--worded findings. Instead, the findings should reflect strengths as well as weaknesses, and even more important, should be structured in a way to provide comparable data across programs, and should be designed so as to provide feedback to aid in the ongoing improvement of program practices. It is regrettable that the tendency to avoid measures which might produce negative results has all but precluded the possibility of learning from project deficiencies (Berman and McLaughlin, 1974).

Given the current developments in the field or research, there appear to be few obstacles to conducting sound, rigorous evaluation of bilingual programs. Such evaluation would reflect the following elements:

(1) Careful collection of meaningful baseline data from selected subjects.

(2) The identification and development of instruments to measure key variables.

(3) The identification of treatment characteristics and documentation on the implementation and development of the program.

(4) The establishment of longitudinality.

(5) The interpretation of results in implementable terms that are meaningful to teachers, policy makers, and researchers.

Whereas the research literature on bilingual schooling is generally lacking rigorious longitudinal evaluations, several such investigations have been conducted in Redwood City (Cohen, 1975) in Culver City (Cohen, 1974), in Montreal (Bruck et al., 1975), and in Illinois (Cohen and Rodríguez-Brown, 1977).

Rationale

Although several attempts have been made to develop comprehensive evaluation studies which will shed some light on the state of affairs of bilingual education programs (Cohen, 1974), most of the studies done, up to now, contain severe problems in the areas of design, usability and applicability of dependent and independent variables, data management, program documentation and interpretation of findings. (i.e. AIR report 1976, Chicago Board of Education, 1976).

It is important then, to review some of the most common problems found in evaluation studies of bilingual education and to describe what could be a realistic design for evaluation of bilingual programs. Realistic in the sense that the author (knowledgeable of the nature and actual functioning of bilingual programs), will try to accomodate all the intrincacies into his/her planning and come up with a design that is valid for the situation although not the strongest from the research point of view. The purpose of the paper is then, to pinpoint some of the issues usually overlooked by people evaluating bilingual education, and to recommend alternative ways to collect, look at, and interpret data.

The Role of Evaluation in the Implementation of Educational Programs

The skepticism with which most people (administrators, teachers, etc.) involved in programs that require annual evaluation reports (i.e. Title I, Title VII) see the role of the evaluator is not unknown to people working on evaluation studies. There is, in general, a misunderstanding as to the role of the evaluator and/or the evaluation studies which need to be clarified.

The role of the evaluator is mainly that of describing the data he/she collects from the program and interpreting those results. The evaluator is not interested in individual teachers and how their students perform. He/she is not interested in looking at individual student scores. His role is mainly to explain, from the data at hand, what the status of the program is. This implies that the evaluator should pinpoint strengths and weaknesses of the program. He/she may directly recommend some changes in the structure and/or design of the program or he/she may recommend that an expert in the discipline involved in the program review the program and change it accordingly. The function of evaluation is to encourage the implementation of better programs.

Of course, the author is conscious of the misuse of misinterpretation of evaluation data. This misuse, though, very seldom comes from the evaluator himself. Most of the time the data is misinterpreted when it gets into the hands of the administrators, teachers and the general public. It is therefore relevant to state that the data is misinterpreted not because the public is against a program (in most cases), but because the public misunderstands the role of evaluation as it is specified above.

After this brief clarification as to the role of the evaluator and the purpose of evaluation studies, a look at the different components involved in the evaluation studies and specifically, bilingual education evaluation

studies, will be covered in the following sections of this paper.

The Evaluator or Evaluation Team

One of the problems found in bilingual education evaluation projects is the fact that the people in charge of the evaluations are either experts in research and evaluation or experts in bilingual education. It is desirable to have a team of people involved in evaluation projects; this way, it is possible to include in it people knowledgeable in research and evaluation as well as specialists in bilingual education.

When the evaluators' expertise is in research and evaluation, they may be not knowledgeable of the intricacies and particularities of the programs studied. The evaluator, as an example, may decide to evaluate bilingual programs by using a strong design that requires a control group. This design supposedly would make the study more valid statistically, but it is almost impossible to find a good control group composed of children from the same linguistic, cultural, and socioeconomic background who are not participating in bilingual programs. Most states that have a large population of linguistic and culturally different children mandate bilingual education and, furthermore, all children who need bilingual education are supposed to participate in the program.

On the other hand, if the evaluation team was formed only by people whose expertise is in bilingual education the evaluation may lack a design at all. What happens in this situation is that tests are administered in great quantities. They may be scored but, then, no one knows what to do with the data. A lot of data mismanagement occurs in this situation. Since there is not a plan for the evaluation, there is no documentation as to how the program was implemented through the year, including such aspects as any environmental changes, personnel changes, etc., which may have affected the results of the testing. A lot of times, means and standard deviations are calculated from the data but no interpretation of results is given.

There is a need then, to have an evaluation team where both, people specializing in research and evaluation and people in bilingual education, work together and compromise on what may be a weak design in terms of research and evaluation, but realistic and feasible in terms of the current issues involved in bilingual education.

Experimental Design and the Evaluation of Bilingual Programs

In the case of bilingual education programs, specifically the programs in those states where bilingual education is mandated, it is almost impossible to talk of a rigorous experimental control random assignment design.

First of all, since most state laws require that schools with a determined number of children of the same non-English or limited English speaking backgrounds have a bilingual program, it would be very hard to find a comparison group formed by children culturally and linguistically similar who do not

participate in a bilingual program.

With this in mind, the evaluator has to look for a design capable of describing the impact, or lack of it, on a program (or treatment) without the use of a control group. Realistically, and taking into consideration the design specifications defined by Campbell and Stanley, (1966), the evaluator, under the restraints provided by the nature of bilingual programs, would have to rely mainly on one of the following designs:

(a) The Before-and-after Design (Center for the Study of Evaluation, 1974). In this design only the experimental (in our case the bilingual program) group is measured. The findings, though, only relate as to the way one programs works.

(b) Time Series Design (Center for the Study of Evaluation, 1974)

(c) A third alternative will be to use participants as their own control. To this end, the district can determine an expected level of attainment described by objectives. This alternative is not a statistical one and may require that the district develop its own criterion-referenced testing system.

In the case of the Before-and-After design, one group of students takes a pre-test, the same group gets the treatment (in this case the participation in the bilingual program) and afterwards, the same group takes a post test. The final results can be compared with norms, if the tests were normed ones. In the case of bilingual programs it is recommended that the districts or the states develop their own norms.

This design can be represented, using Campbell and Stanley's signs as follows:

	Time		
E - group	(pre) O	x	(post) O

Where: E - group = experimental
O = measurement or observation of some kind
pre = pretest
post = post test
X = treatment which should be very well documented in this type of design.

One of the problems with this design is that it is very hard to explain if the results are due mainly to the program or some other factor. It is impossible to determine what the results would have been without the program. A good aspect of the design is that by having to follow only one group, the evaluator can spend more time documenting the program including materials and activities and their relation to the objectives to be attained.

As to reporting scores, the evaluator could report for each test and subject tested; 1) the number of students per school (it should include only

students that were pre- and post tested), 2) the mean score for the pre-test, 3) the mean score for the post test and 4) t-test results. The t-test tests the significance of the difference between the pre- and post-test scores.

Usually, children perform better as they grow older, so you will probably find significant differences between pre- and post test results in areas such as cognitive development and achievement. If no significant difference is found, the evaluator should look into the treatment documentation data to make assumptions on this regard. This way, it may be possible to explain the happening either as a testing error or a real non-progress situation.

The evaluator can examine the results by comparing them to standardized measures. In this case, it will be important for the evaluator to describe the population on which the test was normed and compare both groups. This is to make sure the groups are similar not only culturally and linguistically but contemporarily and socially.

The evaluation report for this design should not be based only on standardized test results (if you have used this type of test and have valid norms). The report should include 1) all the documentation of the program, 2) statistical report of pre- and post test data and t-test, 3) a report on the norm referenced data collected, and, of course, 4) some interpretation as to what the findings may mean and recommendations as to how the program could be improved.

One of the suggestions given to evaluators who have to use this design is to develop some sort of measurement by objectives evaluation (ideally a criterion-referenced system). This way it is possible to report on the program's strong and weak areas according to the performance of students in regard to different objectives. It is recommended, too, that the evaluator choose tests that are sensitive to the grades of the programs being evaluated.

In the Time Series design with an experimental group only, the experimental group is tested several times before and after the treatment (bilingual program) starts and at specific intervals of time.

Figure 2 shows a diagram for this design using Campbell and Stanley's symbols;

Figure 2

	TIME							
	1	2	3	4		5	6	7 8
EXPERIMENTAL GROUP	0	0	0	0	X	0	0	0 0

Where: O = measurements
X = treatment or program

The main steps in the implementation of this design can be described as:

1. Choose tests or measures, reliable, valid and proper to use with the population you are working with, which can be used repeatedly.

2. Choose the composition of the experimental group to be tested (i.e. the same group tested several times, randomly selected groups each time or successive groups of students.)

3. Make sure to collect at least 3 measurements at regular intervals before the program X starts.

4. Check and document the implementation of the program.

5. Collect measurements at the same regular intervals as before the program.

In relation to the composition of the experimental group to be tested, if the same group is tested all the time or if randomly selected samples from the experimental group are selected, the design can be called a longitudinal time series design. If successive groups of students who supposedly represent the group are tested each time, the design is called a successive groups time series. The nomenclature used is the one given by the Center for the Study of Evaluation - UCLA, 1974.

The following are several aspects that the evaluator using this design should take into account when describing in relation to the program implementation.

1. It is important to specify if the program was implemented and when. Implementation data including exact dates should be used in documenting the program.

2. Make sure that documentation information includes any happennings occurring during time of implementation (i.e. a new teacher came in at the same time the program was being implemented). To make any statements related to effect of the program, the influence of other aspects such as bringing in a new teacher should be minimized because a correct explanation of the effect of the program would then be impossible.

3. Changes in the method of collecting data should be documented and explained. Were the same tests or instruments used all the time? If they were different, how different were they? Is there any way to make the scores comparable?

4. Explain any changes in the composition of the experimental group. Is the group the same as when the program started? If different, how different? If you think the group has changed a lot, it would be helpful to look at another set of list scores. For example, if the program treatment received by the experimental group is in reading, you should collect data on math. Later, if there is any question as to whether the nature of the experimental group has changed and produced an effect in scores (i.e. brighter children came into the program), the math as well as a reading tests results could be compared. If the reading scores go up significantly but the math scores do not, it is possible to say that the higher score in reading is due to the program.

5. Check to see if the results show a cyclical pattern. If this is so, then the results are not due to the program. For example, it

may be that results peak at some point during the year. This can be checked by looking at scores from previous years.

Even when a lot of different data is collected to document the implementation of the program, there exists the possibility that the results from the evaluation are not due to the program. Since this design is not very strong, it is called a "quasi-experimental design" (Campbell and Stanley, 1963).

It is the view of the author, that due to the problems involved in finding a control group to evaluate bilingual programs, the next best design to be used is the Time Series Design. It is recommended that either the same group or a randomly selected sample from the experimental group be tested each time so that it is possible to attain some longitudinality with the data collected.

Independent and Dependent Variables and the Evaluation of Bilingual Programs

The variables selected for measurement in any evaluation study are determined by two primary considerations. These are:

- What are the goals or objectives of the school district or educational program?
- What controls are needed to provide meaningful comparisons among programs?

From answers to the latter question, the set of independent variables can be developed. With this idea in mind, a series of independent variables is defined that could be relevant to a longitudinal study of bilingual schooling. These variables are loosely defined and only attempt to touch on factors that should be considered for an evaluation design. The importance of each factor as a source of variation that must be controlled, must be made in light of the specific evaluation settings.

These variables have been divided into three groups or categories.

I. Contextual Variables

II. Student Variables

III. Treatment Variables

A list of these variables under each category is given below.

I. Contextual Variables

A. School district characteristics

1. size
2. resources
3. ethnic composition
4. degree of integration
5. SES composition

B. Community characteristics

1. density
2. SES composition
3. degree of integration
4. occupational make-up
5. political involvement including involvement in school district
6. educational attitudes

C. Parent characteristics

1. schooling
2. occupation
3. ethnicity
4. attitudes
5. involvement
6. SES status
7. dominant language
8. children-home life

II. Student Variables

A. Physical characteristics

1. sex
2. size
3. health (physical handicaps)
4. age

B. Education

1. level of schooling
2. years of schooling
3. schooling characteristics
   a. grades (marks)
   b. continuity
   c. special program (other bilingual programs)
4. attitude toward school and education
5. dominant language
   a. writing
   b. reading
   c. speaking
   d. listening
6. achievement
   a. home language context
   b. English context

C. Peer relations

D. Language association

1. years in U.S.
2. age of first association with English
3. duration and time history with association
4. intensity of association

III. Treatment Variables

A. Setting

1. school characteristics
2. classroom characteristics
3. other programs employed

B. Program characteristics

1. size
2. staffing characteristics
3. personnel relations
4. selection criteria
5. curriculum

a. design
b. organization
c. role of culture

6. materials
7. language usage

a. allocation to subjects
b. ammount
c. method
d. peer usage
e. student teacher usage

Some people have suggested or even used program models as an independent variable (Board of Education, City of Chicago, 1976). It is the author's personal view that program models and instructional structures as they are implemented now are very loosely defined.

Even if you are comparing two bilingual programs described as half-day programs, the two may be structured very differently and may not be providing an equal treatment. It is for this reason, that comparisons between different program models are not accurate and present a lot of measurement and interpretation problems. As one can see, the list of independent variables is lengthy to begin with, and this list is by no means complete.

The answers to the question concerning the goals and objectives of the program should define the dependent variables to be measured. With respect to dependent variables, we try here to cover general areas of concern. As of now, there are not many good tests available for the measurement of common bilingual objectives. We feel that any test chosen should measure the objectives of the

program and should not be used only because of its availability. This may mean the development of tests designed specifically for a given bilingual program.

Some usual measures that reflect bilingual objectives are:

I. Achievement

A. Reading

1. home language
2. English

B. Academic and cognitive ability

1. in home language
2. in English

C. Math

D. Science

E. Social studies

F. Language Ability

1. listening
2. reading
3. writing
4. speaking

II. Language Dominance and Parity

III. Affective Development

A. Self-esteem

B. Self-concept

C. Attitudes

One of the most prominent dilemmas in evaluation is the validity of measures on the independent and dependent variables. Some considerations in this area are:

1. The measures should at least be reliable.

2. When normed scores are used, the norming group should have similar characteristics to the children being measured (i.e., language,

culture, socioeconomic status, etc.), and measurement should be made on the treatment groups at the same time during the school year that the norm group was tested.

3. If the instruments have parallel forms in English and a second language, the forms should have been adapted and not just translated. If translated, they should have been field tested.

4. The language used to obtain measures should only include the language the children use at their developmental level.

5. The measures should be culturally sensitive.

6. Administration and scoring should be straightforward and objective.

Some unique problems of validity occur when pretest and posttest differences or more complex measures of change are used. These problems are aptly described by Lord (1963). Namely, these problems include the regression effect paradox, the reliability of estimated change, the effect of change on group heterogeneity, spurious correlation between change and some other variable.

Another assault on validity of longitudinal studies occurs by the mere fact that the time period over which measures are made is greater than in the one-shot design. Thus, changes can occur in the time-dependent contextual variables. For this reason, these variables should be measured on more than one occasion along with the dependent measures. Ideally, such measures should be made concurrently.

Data Collection and Management in Evaluation Projects

One of the most straightforward tasks of evaluation studies in general and particularly longitudinal studies is data collection and management. Yet, this talk is usually the one that requires the most effort and is usually poorly done, resulting in invalid evaluations. Competency of data gatherers and their managers is mandatory. Some considerations in collecting and managing the data are:

1. Data should be maintained on a per student basis.

2. All students should be given one and only one unique identification number and this should be recorded on all information collected.

3. A computerized data base should be developed where possible to organize and maintain the data.

4. Sorting of students by informative identification numbers can provide an easy to use directory.

5. Meaningful identification numbers can be produced by using indicators of student characteristics such as the school, program, and section he is enrolled in, his birth year, year he entered the program, and grade he entered the program, etc.

6. Computer routines for validity checking should be incorporated into the data management system.

7. Simple editing, sorting, and merging routines should be set up for production use.

8. All data colection and management activities should be the responsibility of one person. This will avoid confusion and misinformation that normally occur when many data gathering activities are undertaken.

9. Many of the data management duties require the technical expertise of a good computer services staff which has some knowledge of statistical software that may be applied for evaluation.

Longitudinal designs are more susceptible to missing data problems through attrition and other reasons. All efforts should be made to avoid missing data. Where such problems do occur, there is very little elegant recourse. Some possible compensation steps which are not without bias are:

- Exclude records that have missing data.

- Estimate missing data from regression equations developed from available data. (In this case, the 95% confidence intervals could be used rather than the point estimates and appropriate maximum likelihood regression techniques could be applied to handle the mixed data forms, that is point and interval values.

- Scale down the evaluation to include only that set of variables for which complete data are available.

As stated, each of these approaches are biased. The degree to which they can be applied depend on the data at hand.

Another primary dilemma of longitudinal evaluation and specifically bilingual evaluation is the comparability of measurement instruments over time. Tests in bilingual education that represent a continuum over various educational and developmental levels are scarce. These tests are usually not normed and thus one level is not related to others. The concept of grade equivalents has not been applied to bilingual measures. Thus measures of student progress over time may have to be developed before meaningful trend analysis can be performed. This is a major problem since a great deal of time, effort, and expertise must be employed to develop tests that measure the same concepts at various levels. The authors have no good suggestions for handling this problem other than to start from scratch.

Conclusion

There is a great need to have good evaluation studies in regard to bilingual education programs. We want to know what the reality in these programs is

today; specifically we want to know what their strengths and weaknesses are. This way, programs can be better implemented.

There are many issues which should be taken into account when designing evaluation studies in bilingual settings. This paper has tried to make the reader conscious of two salient issues related to evaluation of bilingual programs: 1) there is a need to identify an evaluation team with diversified interests and expertise and, 2) the evaluator must be aware of the restrictions in properly identifying a control group.

Under the constraints caused by the nature of bilingual education programs, it seems that the time series design, where a random sample or the whole population is tested at different points of time, is the best alternative to strong designs used in evaluation studies. Since this design is a weak one and there is not a control group involved, the need for a good, comprehensive documentation of the implementation of the program is crucial. This will help evaluators clarify and explain whether the findings from the evaluation study can be interpreted as due to the program and not related to other variables unrelated to the program.

The author recommends that independent variables used in the evaluation studies should be chosen in relation to the particular needs and characteristics of the setting being studied. It is recommended, though, that comparisons among children attending different program models (i.e. half-day programs vs. self-contained) would not be made at this point of time. It is not possible to compare children attending one program model against another because the program attended may not be the same even within a school year. Program models, too, have been very vague and ill-defined up to now. Although two programs may be called the same name (i.e. half-day), they may provide completely different treatments and as such they can not or may not be compared as to their effect on student achievement variables.

Finally, the need for a data management system is stressed, especially in longitudinal studies. This is a component of evaluation which is often overlooked but which can produce problems when analysing data and drawing conclusions with findings.

REFERENCES

American Institute for Research in the Behavioral Sciences. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs. Palo Alto, California, 1977 - ERIC number FL-008-505.

Berman, P., & McLaughlin, M.W., Federal Programs Supporting Educational Change, vol. 1: A model of educational change. prepared for the U.S. Office of Education, H.E.W., R-1589/1-HEW, Santa Monica, CA.: Rand Corporation, 1974.

Bruck, M., Lambert, W.E., & Tucker, G.R.,"Cognitive Consequences of Bilingual Schooling: The St. Lambert Project Through Grade 8." Paper presented at the 9th Annual TESOL Convention, Los Angeles, March 4-9, 1975.

Campbell, Donald T.and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Rand McNally and Co., Chicago. 1966.

Center for the Study of Evaluation. How to Calculate Statistics and Also Understand Them. UCLA, Graduate School of Education. Los Angeles, California. 1974.

_____. How to Select a Design, Display and Interpret Results. UCLA, Graduate School of Education, Los Angeles, California. 1974.

Chicago Board of Education. Final Evaluation Report of State-Supported Bilingual Education Program for Fiscal 1974. Department of Government Funded Programs, Chicago Board of Education, City of Chicago, 1974.

_____. Final Evaluation Report of State-Supported Bilingual Education Programs for Fiscal Year 1976. Department of Research and Evaluation. Chicago Board of Education, City of Chicago, 1977.

Cicourel, A. et al., Language Use and School Performance. N.Y. Academic Press, 1974.

Cohen, A.D. "The Culver City Spanish Immersion Program: The first two years." The Modern Language Journal, 1974.

Cohen, A.D. A Socialinguistic Approach to Bilingual Education: Experiments in the American Southwest. Rowley, Mass.: Newbury House, 1975

Cohen, A.D., and Rodríguez-Brown, Flora. "Evaluation in Moderate to Small School Districts: Downstate Illinois Manuscript." Bilingual Education Service Center. Arlington Heights, Illinois. 1977.

Cohen, A.D., & Laosa, L.M.,"Approaches to Second Language Instruction: A Research Model." Paper presented at the American Educational Research Association Meeting, Washington, D.C., March 31-April 3, 1975.

Epstein, Noel. Language, Ethnicity and the Schools. Policy Alternatives for Bilingual-Bicultural Education. Institute for Educational Leadership: The Geroge Washington University, Washington, D.C., 1977.

Lord, F.M., Elementary Models for Change. in C.W. Harris (ed.), Problems in Measuring Change. Madison: University of Wisconsin Press, 1963.

Ramirez III, M. et al. Spanish-English Bilingual Education in the United States: Current Issues, Resources and Recommended Funding Priorities for Research. National Institute of Education Contract #N1E-C-74-0151. East Santa Cruz, California: Systems and Evaluations in Education, 1975.

Saville, M.R. & Troike, R.C. A Handbook of Bilingual Education. Revised edition, Washington, D.C.: Teachers of English to Speakers of Other Languages, 1971.

Swanson, María. "Bilingual Education: The National Perspective." In Responding to New Realities. Edited by Gilbert A. Jarvis. The ACTFL Review of Foreign Language Education #5. National Textbook Co., Skokie, Illinois. 1974

Troike, R.C. "International Perspective in Bilingual Bicultural Education." Annual International Conference on Bilingual Bicultural Education, New York City, N.Y., May 15-18, 1974.

Tucker, G.R. & d'Anglejan, A. "Some Thoughts Concerning Bilingual Education Programs." The Modern Language Journal, 1971, 55.8, 491-493.

U.S. Commission on Civil Rights. A Better Chance to Learn: Bilingual Bicultural Education. U.S.C.C.R. Clearinghouse Publication 51, May, 1975.